Building Detection from Monocular VHR Images by Integrated Urban Area Knowledge

Andrea Manno-Kovács, *Member, IEEE,* Ali Ozgun Ok, *Member, IEEE*

***Abstract*—This paper proposes an approach for building detection from single, very-high-resolution (VHR) optical satellite images by fusing the knowledge of shadow and urban area information. One of the main contributions of this work is in the integration of urban area information: unlike previous works, we use such information to substantially revise and improve the initial shadow mask. Additionally, we present an effective way to discriminate dark regions from cast shadows, a task that has continuously been reported to be very difficult. In this study, we benefit from graph-cuts to produce a comprehensive solution for automatic building detection: a flexible multi-label partitioning procedure is proposed, in which the number of optimized classes is automatically selected according to the contents of a scene of interest. The results of the evaluation of 14 demanding test patches confirm the technical merit of the proposed approach, and its superiority over three recently developed state-of-the-art methods.**

***Index Terms*—Automated building detection, urban area detection, graph-cuts, flexible multi-label partitioning, satellite images.**

I. Introduction

BUILDINGS, while being one of the easier object classes for automatic detection, present a major challenge when they need to be detected in monocular images.

Lately, various building detection approaches based on local features have been introduced: four different local feature descriptors were tested and two fusion methods were introduced in [1] to improve the results of detection; a graph-based strategy was developed in [2] utilizing corners and edges of buildings; a Marked Point Process framework was built in [3], [4] using lower and higher level features to find the locations of buildings; an automatic approach was introduced in [5] combining the strength of energy-based approaches with the distinctiveness of corner features. In [6] it was shown that shadow information can be efficiently used for building localization, introducing a new fuzzy landscape strategy to uncover the relation between buildings and shadows and using iterative graph-cuts for automated building detection. In [7] an original approach was proposed based on shadow information and graph-cuts, utilizing multi-label graph partitioning for building detection in an unsupervised manner. More recently, a higher order conditional random field was integrated for building detection [8], in which an unsupervised segmentation strategy was also utilized for shadow detection.

A. Manno-Kovács is with Distributed Events Analysis Research Laboratory, Institute for Computer Science and Control, MTA SZTAKI, Hungarian Academy of Sciences, Budapest, Hungary. E-mail: andrea.manno-kovacs@sztaki.mta.hu;

A. O. Ok is with the Department of Geodesy and Photogrammetry, Nevsehir H.B.V. University, 50300, Nevsehir, Turkey. E-mail: ozgunok@nevsehir.edu.tr

In certain cases, a coarse classification of images into urban and non-urban areas is the only option. Harris and SUSAN corner detectors were analyzed in [9] and found to be efficient for distinguishing between different structures in a scene (urban / non-urban). In [10] the scale-invariant feature transform (SIFT) was combined with graph theory for the joint detection of urban areas and buildings. Gabor feature points were extracted in [11], followed by a voting matrix construction to represent the probability of urban areas. [12] showed that the Modified Harris for Edges and Corners (MHEC) method [13] is able to represent urban areas more precisely; moreover, the main direction of the gradients in the close proximity of feature points is an important feature during the detection of urban (i. e. built-in) areas.

The method presented in this letter aims to detect buildings from single optical VHR satellite images by integrating the information of built-in areas, as this information may help to achieve more accurate detection in complex and challenging cases. Handling urban area and separate buildings jointly and combining detection results has been introduced in [10], where SIFT keypoints were utilized both for urban area and building extraction purposes. However, the approach is designed to handle relatively non-dense urban areas. The same authors proposed a sequential combination of both approaches in [1], first detecting urban areas, then searching for buildings within such regions. However, because of the sequential process, the performance of the building detection will always be limited by the performance of the urban area detector.

In this letter we introduce a novel framework for building detection in monocular VHR images, based on knowledge of urban environments, using graph-cuts (Fig. 1). We assess the presented approach on 14 test patches of two different space-borne sensors (QuickBird and IKONOS-2), and the results prove the quality and reliability of the framework proposed for the detection of buildings. We also made comparison with three other state-of-the-art approaches to confirm the proposed method's superiority.

The structure of this paper is as follows: we state the contributions of this study in Section II, we describe the method in detail in Section III followed by the description of the dataset and parameter analysis in Section IV, Section V presents the experimental evaluation results and we conclude with Section VI.

II. Contributions

The main contribution of this paper is developing a building detection approach that takes advantage of information of

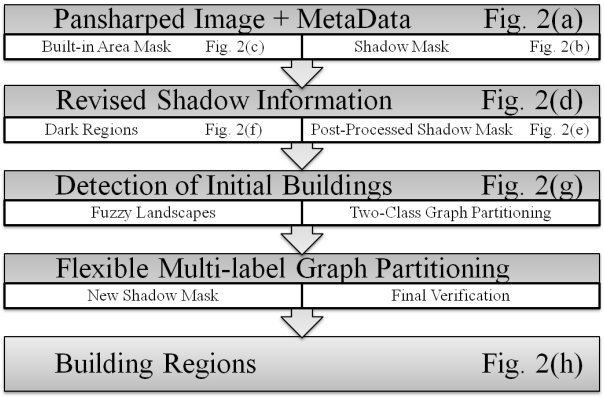

Fig. 1. Diagram of the proposed method

urban areas in an embedded manner. Unlike previous methods that search for buildings within image regions labeled as built-in areas, therefore limited by the performance of the applied built-in area detectors, our approach integrates the information of urban areas and solar angles to revise and improve the shadow mask. With this step, we are able to not only eliminate improperly detected shadow regions, but we can also handle the unbalanced distribution of built-in / not built-in areas (e.g., large water surfaces in the image). The latter in turn provides the opportunity to automatically identify and separate dark, but non-shadow regions from cast shadows. In this sense, we propose a flexible multi-label partitioning procedure in which the number of classes to be optimized is automatically defined by the contents of the scene of interest. The dark regions are defined as a separate class before the partitioning, which makes their separation from shadows more efficient after the partitioning step, resulting in better building region extraction and a better overall performance.

III. Proposed method

The proposed approach uses orthorectified multi-spectral images (Fig. 1). Metadata[1] including the Sun angle at the time of the acquisition, is attached to the image. Firstly, two masks, a vegetation mask and an initial shadow mask, are extracted. Then, the Modified Harris for Edges and Corners (MHEC) approach is applied to detect urban areas, which is based on the modification of the Harris detector's characteristic function, which describes urban areas more accurately than other point detectors [12]. With the aid of the solar information and the built-in area mask, the shadow mask is evolved and fuzzy landscapes are generated. The approach consists of two levels of graph partitioning to detect building regions. The first level uses iterative graph-cuts to recognize initial building regions, whereas the second level is devoted to multi-label graph optimization to finally distinguish building regions. The final verification step aims to identify and label building regions, whose directional neighborhood reveals shadow evidence. Some of these stages are already described in detail in [7] and [12], therefore, we only concentrate on new contributions in the following sub-sections.

[1]The solar angles (azimuth and elevation [6]) are assumed to be fixed during the period of image acquisition.

A. Revised shadow information

To identify the initial shadow map (M_{IS}), first we apply a shadow detection technique presented in [14], which computes a ratio map using saturation and intensity components of the Hue-Saturation-Intensity (HSI) space. This step is followed by the well-known Normalized Differential Vegetation Index (NDVI) calculation to subtract regions belonging to the vegetation cover and finally, a constrained region-growth process is applied (see [7] for details). The M_{IS} map is then revised based on the estimated built-in area map (M_{BA}), which is calculated by an orientation selective voting matrix based on MHEC feature points. The effects of the M_{BA} map, by testing weight spreading parameters in the voting matrix (see [12] for details), is shown in Table I.

The built-in area mask M_{BA} is used for two purposes: (i) revising the shadow mask M_{IS}, and (ii) excluding dark non-shadow evidence in M_{IS} from true shadow evidence. The former helps to better detect cast shadows in a scene of interest, wherein a large portion is covered by a region whose spectral reflectance resembles shadow regions. A typical example is given in Fig. 2, where a large water body is visible in a scene. In such a case, the M_{IS} map (Fig. 2(b)) is poor for the cast shadows of buildings, due to the spectral similarities between the water surface and shadows. In this study, the built-in area information (Fig. 2(c)) is used to mitigate these problems. First, we search for connected shadow components (8-neighborhood connectivity) in the shadow mask M_{IS}, and overlay each component with the built-in area mask M_{BA}: if more than half of a component is found to be outside of the built-in area, we discard all pixels of that component from the precomputed HSI-based ratio map and regenerate the revised shadow map M_{RS} (Fig. 2(d)). After this step, the detected shadows cast by buildings are significantly improved in M_{RS} compared to the M_{IS} mask. However, the revised shadow mask may still incorrectly label dark surfaces as shadow. Therefore, once again, we overlay each shadow component ver the built-in area mask M_{BA} and only accept shadow components with more than half of the pixels in the built-in area. Finally, as a post-processing step, we perform a morphological opening over each shadow component, using a specific directional flat structuring element, generated from the known solar information [7]. The goal of this latter step is to eliminate cast shadows of short objects (e.g., cars, garden walls and fences, etc.) to obtain the final shadow mask M_{PS} (Fig. 2(e)).

B. Detection of dark regions

The other purpose of the built-in area mask is to detect non-shadow, dark regions. Therefore, components rejected from the shadow mask M_{RS} are supplied to generate a new mask that represents dark regions (M_{DR}) (Fig. 2(f)). After this step, if the evidence in mask M_{BA} defines the built-in areas perfectly, we expect to have an exact separation between the two cases (cast shadow vs. dark region). Unfortunately, this is rarely possible because of the problems associated with the detection of built-in areas (for example, see the separated building on the small island in the center of Fig. 2(a)). To compensate

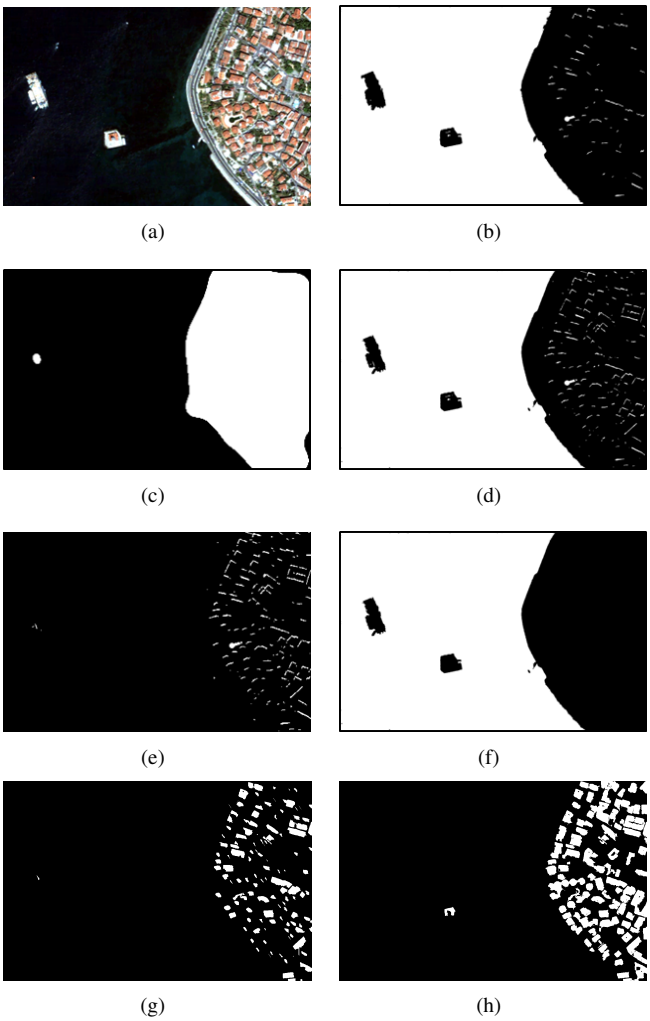

Fig. 2. Main pre-processing steps: (a) Test patch #12 (RGB); (b) initial shadow map M_{IS}; (c) detected built-in area map M_{BA}; (d) revised shadow map M_{RS}; (e) post-processed shadow map M_{PS}; (f) dark regions detected M_{DR}; (g) initial building regions; (h) final building regions.

for these drawbacks, we propose to use solar information to evaluate the regions belonging to dark regions in the mask M_{DR}. As an improvement over [7], we expand the approach to impose a maximum height threshold (T_{maxH}) for objects that cast shadows. We generate a flat structuring element ($\nu_{L,\lambda+\pi}$) that maintains the solar azimuth angle A and the directional information λ ($\lambda = A - \pi/2$) with a minimally connected single edge segment L, where the length of the edge segment (l) in the image space can be computed on the condition that the surface on which shadows fall is flat:

$$l = \left\lceil \frac{T_{maxH}}{\tan\phi \;.\; c} \right\rceil, \tag{1}$$

where ϕ denotes the solar elevation angle, c denotes the resolution of the image, and the operator $\lceil . \rceil$ rounds to the next-larger integer. Once the structuring element is generated, we label each component in M_{DR} with an 8-neighborhood connectivity analysis and apply morphological erosion to each component independently. We keep track of the components that do not entirely disappear after the erosion, and finally label all pixels of those components as dark region (Fig. 2(f)).

C. Fuzzy landscapes and detection of initial building regions

Given a shadow object and a non-flat line-based structuring element, the fuzzy landscape around the shadow object along a given direction can be defined as a fuzzy set of membership values in image space. The generated landscape for each shadow object provides an indication of a surface discontinuity near the shadow region (e.g. buildings, walls, vegetation canopies etc.). Therefore, an automated fuzzy pruning step [6] is applied to consider only the landscapes belonging to building regions.

In this stage, we consider the building detection task as a two-class partitioning problem where a given building region has to be separated from the background (building vs. others). A region-of-interest (ROI) is determined for each shadow object, and pixels corresponding to foreground/building (T_F), background/non-building (T_B), and unknown classes (T_U) in the ROI are automatically labeled using shadow object and fuzzy landscapes [6] (Fig. 2(g)). To solve the partitioning, we model T_F and T_B using Gaussian Mixture Models (GMMs) and use the GrabCut approach [15] with the following energy formula:

$$E(f) = \sum_{n \in P} D_n(f_n) + \gamma_1 \sum_{(m,n) \in N} e^{-\beta \|z_m - z_n\|^2}, \tag{2}$$

where $D_n(f_n)$ favors the label preferences (foreground/background) for each pixel n based on the observed pixel values z and eventually depend on the results of mixture modeling. N defines the set of neighboring pixel pairs computed in the 8-neighborhood, and β and γ_1 are the constants that determine the degree of smoothness, where $\beta = (2\left\langle \|z_m - z_n\|^2 \right\rangle)^{-1}$ with $\langle . \rangle$ expectation operator. To complete the partitioning and to estimate the final labels of all pixels in the bounding box, a minimum-cut/max-flow algorithm is applied.

D. Flexible multi-label graph partitioning and verification of building regions

The goal at this stage is to uncover final building locations using a multi-label graph partitioning strategy initialized by the information collected in previous stages. For a similar purpose, a rigid four-label energy minimization was proposed in [7]. However, unlike in that work, we have to take into account the diversity of urban areas and the environmental-illumination conditions during imaging. In this respect, the classes *vegetation* and *dark region* can be considered optional, and their presence principally depends on the contents of a scene and the season/date of imaging (assuming that there are buildings in the images and shadows cast by buildings are not entirely occluded). Therefore, in this study, we propose a flexible multi-label partitioning strategy with at most five classes: *building*, *vegetation*, *shadow*, *dark region* and *others*. Given the existence of the classes *vegetation* and *dark region*, the number of classes in the multi-label optimization are adopted automatically (varying between three and five). Thus, given a set of pixels $\mathbf{z} = (z_1, z_2, \ldots, z_N)$, and a set of class labels $L \in \{1, \ldots, l\}$ where l depends conditionally on the number of observed classes, our aim is to find the optimal mapping

γ_1	Precision	Recall
1	85.8	74.1
5	**83.5**	**84.4**
10	82.5	83.1
20	81.2	76.7

$\sigma_{i,x}$	$\sigma_{i,y}$	σ_i	Precision	Recall
1	2	1	82.1	85
2	4	2	83.2	84.9
3	**6**	**3**	**83.5**	**84.4**
4	8	4	82.7	84.3

TABLE I
PARAMETER TESTS FOR γ_1 AND M_{BA} WEIGHT SPREADS (SEE [12]).

from the data $\mathbf{z}$ to the class labels L. We utilize the energy function in Eq. 2 and initialize a GMM for each observed class. To minimize the energy $E(f)$ for multi-label optimization using graph cuts, we use the standard α-expansion move algorithm [16], [17], [18]. After the partitioning, we extract the regions labeled as *building* from the results of the optimization. For verification, we first extract the shadow class after the multi-label partitioning, and generate a new shadow mask M_{NS}. Then, we remove building regions which do not have any shadow evidence around their directional neighborhoods. Finally, we employ an area threshold to remove small artifacts mislabeled as *building* (Fig. 2(h)).

IV. DATASET AND PARAMETERS

The test patches used in this study consist of 14 originally orthorectified and pansharpened images selected from two different well-known and well-issued VHR satellite sensors, IKONOS-2 (1 m), and QuickBird (0.60 m). The used imagery includes four multi-spectral bands (R, G, B and NIR) with a radiometric resolution of 11 bits per band. All test patches are purposefully selected to uncover the potential of the proposed approach, emphasizing weaknesses of state-of-the art methods and including different urban area characteristics at the same time. Additionally, patches with a variety of dark regions are chosen, e.g., pools, lakes, sea surface, roads, burnt areas, as well as cast shadows due to topographic discontinuities.

The already investigated parameters from [7] are kept the same. The results for testing γ_1 of flexible multi-label partitioning and weight spreading parameters $\sigma_{i,x}$, $\sigma_{i,y}$ and σ_i for calculating M_{BA} map are presented in Table I, showing the overall precision and recall rates (bold text marks the applied settings). As it is shown, performance may slightly decrease with other settings, however, the algorithm is not very sensitive to the parameters.

In fact, there are only two new parameters to be set, the maximum height threshold for buildings (T_{maxH}) and the number of GMM components for class *dark region*. In this study, the threshold T_{maxH} is fixed to 50 m, which is a practical value for buildings. For the number of GMM components of class *dark region*, we use the same value as for shadow regions, setting it to 2, because of the similar radiometric characteristics.

V. EXPERIMENTS

The reference data consisting of building regions were produced manually by a qualified human operator. To evaluate the pixel-level performance, we use the well-known P precision, R recall and F-measure values.

Previous approaches like [6] and [7] suffer from errors caused by newly installed asphalt pavements, water bodies, burnt areas or simply discontinuities, which are incorrectly detected as shadows in the preprocessing step (examples are shown in Figure 3). These errors result in large false positive areas, reducing detection accuracy significantly. With our introduced contributions, misdetections caused by such problems are successfully mitigated, improving detection results. The MPP [3] method assumes rectangular buildings based on color, shadow and gradient evidences, and the absence of these evidences has crucial effects on detection accuracy. Therefore, weak color evidence (grey rooftops) or varying building shapes (curved structures) cause poor detection results. However, this method copes with asphalt pavements and water bodies successfully.

Quantitative pixel-level detection results are presented in Table II, where the method is also compared to three state-of-the-art approaches MPP [3], GrabCut [6] and Multi-label Partitioning [7]. Test patches $\#4$ and $\#13$ are shown row-by-row in Figure 3 along with the pixel-level detection results: the colours green, red and blue represent TP (true positive), FP (false positive) and FN (false negative) pixels, respectively. Detection results of all test patches indicate that most of the building regions are correctly detected due to the proposed contributions. At the pixel-level (shown in Table II) an overall F-score is computed to be almost 84%. Thus, despite the complexities and difficulties in the test images, the results are encouraging and prove the proposed method's viability.

VI. CONCLUSION

In this letter, the built-in area evidence is exploited to revise and post-process the shadow mask in a unique fashion. The evidence allows us to detach regions belonging to dark objects from cast shadows with the aid of the solar information using a flexible multi-label graph partitioning strategy. The proposed approach is tested for 14 image patches of two optical VHR sensors, and the evaluation indicates that the proposed approach achieves quite promising results and outperforms three state-of-the-art methods chosen for comparison.

In the future, we plan to focus more on buildings whose cast shadows are completely occluded, and on cases where snow or haze affects the detection. Novel features could be integrated to solve these challenges: by exploiting the orientation information of the feature points, edges from such cases could be investigated. Also, the probabilistic results after fusing the voting matrices could be directly integrated into graph partitioning, which might result in further improvements. The present method cannot discriminate cloud and terrain shadows from building shadows, therefore worldwide available DSM data may be integrated in the framework. Concerning the increasing ground sampling distances (GSD) of images, we plan to test our approach on WorldView-3 images with 30 cm GSD. However, it might not be possible to obtain complete shadow regions with images having significantly better GSD (e.g. 10 cm).

ACKNOWLEDGMENTS

The authors would like to thank Dr. Csaba Benedek for providing the results of his approach for comparison; and for the comments of the two anonymous reviewers which helped to improve the paper. This work was partially supported by the ESA DUSIREF project.

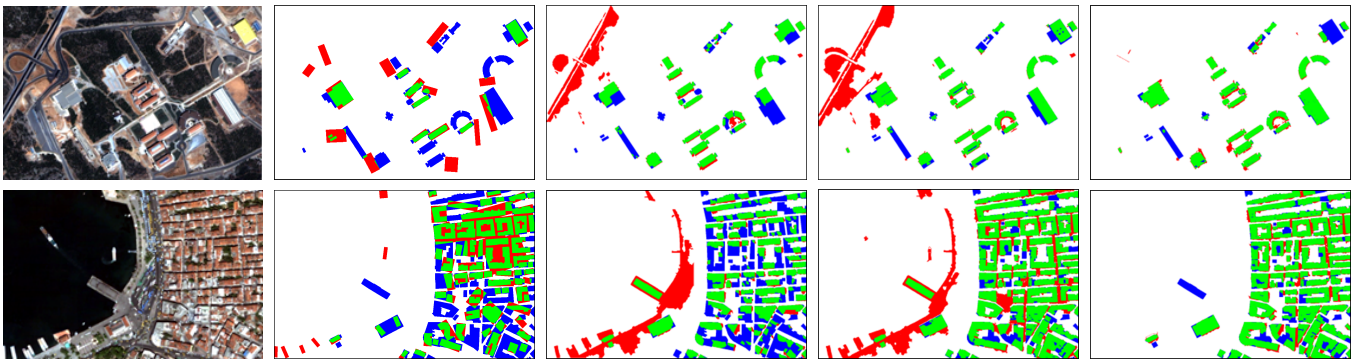

Fig. 3. Results of building detection for test patches #4 and #13. (first column) Test patches, (second column) the results achieved by the MPP [3], (third column) the results achieved by the GrabCut [6], (fourth column) the results achieved by the Multi-label Partitioning [7], and (fifth column) the results achieved by the proposed approach. The colours green, red and blue represent TP, FP and FN pixels, respectively. The entire dataset and their results are provided in the following website: http://biz.nevsehir.edu.tr/ozgunok/en/378.

Database	Pixel-based Performance (%)											
	MPP [3]			GrabCut [6]			Multi-label Partitioning [7]			Proposed method		
Test Image (size)	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
#1(560 × 367)	35.1	48.8	40.8	59.1	58.6	58.8	36.5	56.8	44.4	**81.2**	**75.0**	**78.1**
#2(554 × 483)	41.0	75.7	53.2	70.8	49.8	58.5	**76.8**	78.9	77.8	74.3	**86.4**	**79.9**
#3(468 × 304)	40.3	63.7	49.4	60.4	76.3	67.4	60.1	**90.2**	72.1	**69.2**	89.0	**77.9**
#4(896 × 600)	39.2	36.1	37.6	54.6	64.8	59.3	52.4	76.7	62.3	**86.6**	**78.8**	**82.5**
#5(1213 × 958)	46.2	82.2	59.2	71.5	61.7	66.2	70.2	**89.5**	78.7	**91.0**	88.1	**89.6**
#6(922 × 634)	41.9	59.2	49.1	46.3	**80.0**	58.7	23.8	74.4	36.1	**87.4**	68.2	**76.7**
#7(928 × 639)	56.1	69.8	62.2	77.5	83.2	80.3	77.2	87.3	81.9	**81.7**	**88.8**	**85.1**
#8(1009 × 695)	49.9	49.7	49.8	72.2	69.4	70.8	68.1	**86.9**	76.4	**86.4**	83.6	**85.0**
#9(1615 × 1209)	37.8	67.7	48.5	47.4	62.3	53.9	40.6	74.6	52.6	**89.9**	**90.2**	**90.0**
#10(1656 × 1240)	32.3	43.7	37.1	30.6	71.5	42.8	20.0	71.4	31.3	**61.0**	**73.0**	**66.4**
#11(1222 × 915)	63.9	53.7	58.4	70.1	92.2	79.6	77.9	**95.9**	**86.0**	**83.7**	87.0	85.3
#12(1311 × 848)	70.8	56.7	62.9	46.5	17.3	25.2	41.1	32.2	36.1	**84.4**	**81.0**	**82.7**
#13(1193 × 772)	67.9	60.2	63.9	62.6	52.3	57.0	67.6	**86.0**	75.7	**86.2**	85.1	**85.6**
#14(1193 × 771)	76.7	66.5	71.2	61.1	43.1	50.5	66.6	71.3	68.8	**84.3**	**85.9**	**85.1**
Average	52.7	59.9	56.1	57.5	61.9	59.6	53.1	78.1	63.2	**83.5**	**84.4**	**83.9**

TABLE II
PIXEL-LEVEL QUANTITATIVE RESULTS FOR MPP [3], GRABCUT [6], MULTI-LABEL PARTITIONING [7] AND THE PROPOSED METHOD.

REFERENCES

[1] B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 211–221, 2011.

[2] S. Cui, Q. Yan, and P. Reinartz, "Graph search and its application in building extraction from high resolution remote sensing imagery," in *Search Algorithms and Applications*. Nashat Mansour (Ed.), InTech, 1991.

[3] C. Benedek, X. Descombes, and J. Zerubia, "Building detection in a single remotely sensed image with a point process of rectangles," in *International Conference on Pattern Recognition (ICPR)*, 2010, pp. 1417–1420.

[4] ——, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 33–50, 2012.

[5] M. Cote and P. Saeedi, "Automatic rooftop extraction in nadir aerial imagery of suburban regions using corners and variational level set evolution," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 1, pp. 313–328, Jan 2013.

[6] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, 2013.

[7] A. O. Ok, "Automated detection of buildings from single vhr multispectral images using shadow information and graph cuts," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 86, pp. 21–40, 2013.

[8] E. Li, J. Femiani, S. Xu, X. Zhang, and P. Wonka, "Robust rooftop extraction from visible band images using higher order crf," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 8, pp. 4483–4495, 2015.

[9] L. Martinez-Fonte, S. Gautama, W. Philips, and W. Goeman, "Evaluating corner detectors for the extraction of man-made structures in urban areas," in *Geoscience and Remote Sensing Symposium, IEEE International*, 2005, pp. 237–240.

[10] B. Sirmaçek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009.

[11] B. Sirmaçek and C. Ünsalan, "Urban area detection using local feature points and spatial voting," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 146–150, 2010.

[12] A. Kovacs and T. Sziranyi, "Improved Harris feature point set for orientation sensitive urban area detection in aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 796–800, 2013.

[13] A. Kovács and T. Szirányi, "Harris function based active contour external force for image segmentation," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1180–1187, 2012.

[14] M. Teke, E. Başeski, A. Ö. Ok, B. Yüksel, and Ç. Şenaras, "Multispectral false color shadow detection," in *Photogrammetric Image Analysis*. Springer, 2011, pp. 109–119.

[15] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 309–314.

[16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[17] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 9, pp. 1124–1137, 2004.

[18] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, 2004.